STATISTICAL GENETIC CONSIDERATIONS FOR MAINTAINING GERMPLASM COLLECTIONS

J. Crossa,[1] C.M. Hernandez,[2] P. Bretting[3], S. Eberhart,[4] S. Taba[1]

[1] International Maize and Wheat Improvement Center (CIMMYT), Lisboa 27, Apdo. Postal 6-641, Mexico D.F., MExico.

[2] Universidad de Colima, Colima, Apdo. Postal 22, Mexico.

[3] Plant Introduction Station, Iowa State University, Ames Iowa 50011, USA.

[4] National Seed Storage Laboratory, USDA, ARS, Fort Collins, Colorado 80523, USA.

ABSTRACT

In regeneration the aim is to maintain at least one copy of each allele present in the original population. Population genetic diversity depends on the number and frequency of alleles across all loci. The objectives of this study for outbreeding crops are: 1) to use probability modeles to determine optimal sample sizes for regeneration for a number of alleles at independent loci; 2) to examine some theoretical problems in choosing core subsets of a collection. Assuming that k-1 alleles occur at an identical low frequency of p_o and that the k^{th} allele occurs at a frequency of $1-[(k-1)p_o]$, for loci with two, three, or four alleles, each at p_o of 5%, 89 to 110 additional individuals are required to retain at least one allele at each of the 10 loci with 90% probability; if 100 loci are involved, 134 to 155 individuals are required. For two, three, or four alleles, when p_o is 3% in each of 10 loci, the sample size required to include at least one of the allele from each class in each locus is of 150 to 186 individuals; if 100 loci are involved, 75 more indivudals are required in the sample. Sampel sizes of 160-210 plants are required to capture alleles at frequencies of 5% or higher in each of 150 loci with 90-95% probability. For rare alleles widespread throughout the collection, most of the alleles with frequencies of 3% and 5% per locus will be included in a core subset of 25 to 100 accessions.

Key word: Genetic resources conservation-Sample size-Allele frequency-Probability models-Core subsets

Genetic resources managers strive to maintain gene diversity during regeneration by attempting to retain at least one copy of each allele present in the original population. Crossa (1989) pointed out that the effectiveness of regeneration for maintaining the gene diversity is related to proper sampling procedures, random genetic drift due to sample size, and optimum seed viability. When sample sizes are large, regeneration is difficult and expensive. Small sample sizes may result in loss of alleles present at low frequency by random genetic drift.

Population genetic diversity depends on the number and frequency of all alleles across all loci, plus the population genetic structure. Marshall and Brown (1975) suggested that the most important measure of genetic diversity is the average number of alleles per locus. Weir (1990) defined genic diversity at a single locus as one minus the sum of squares of the allelic frequencies. For outbreeding species, gene diversity and the proportion of heterozygosity are equivalent. In contrast, self pollinating species may have much gene diversity among accessions but few heterozygous individuals within accessions.

The concept of a core collection was introduced by Frankel and Brown (1984) and Brown (1989a,b) with the idea of minimizing the costs of germplasm conservation while insuring maximum genetic diversity in a collection. Later, the authors described how to form a core subset using information on the origin and agronomic and morphological

characterists of the accessions. When forming a core subset curators must answer two important questions: 1) what is the optimum number of accessions for retaining most of the alleles present in a given collection and 2) how to select accessions for the core subset.

The objectives of the present study for outbreeding crops are: 1) to use probability models that incorporate the number of alleles at independent loci to determine optimal sampling procedures and sizes for regenerating germplasm accessions; 2) to assess some theoretical problems in choosing core subsets of a collection.

BINOMIAL AND MULTINOMIAL PROBABILITY MODELS FOR DETERMINING OPTIMAL SAMPLE SIZES FOR REGENERATING GERMPLASM ACCESSIONS

Consider a random mating population of infinite size and in Hardy-Weinberg equilibrium that can be subdivided into many highly homozygous lines. Assume that the organisms are diploid and that there are two classes of alleles per locus (a_1 and a_2), a_1 occurs with frequency p_1 and a_2 with frequency p_2 ($p_2=1-p_1$). Sampling, at random, individuals of n different lines is equivalent to drawing, at random, one gamete after another n times from the original random mating population. This represents n independent, repeated Bernoulli trials. Therefore, the number of lines with a particular allele in a sample size n, is a random variable

with a binomial distribution. The probability of including x_i alleles of class a_1 in a random sample of size n is

$$P = \binom{n}{x_i} p_1^{xi} p_2^{n-xi} \quad \text{(Eq. 1)}$$

Then the probability of including in the sample at least one allele of class a_1 is $P=1-p_2^n$ (Eq. 2).

For k (k>2) classes of alleles, a_1, a_2, ..., a_k of a locus with frequencies of p_1, p_2, ..., p_k, the number of lines in a sample of size n with certain number of alleles from each allele class in a sample of size n, is a random variable with a multinomial distribution. This case represents independent, repeated trials that generalize from Bernoulli trials with two outcomes to trials with more than two outcomes. Therefore, the probability of obtaining each allele class x_i times in a random sample of size n is

$$P = [n!\ \Pi^k_{i=1} (p_i)^{xi}]/[\Pi^k_{i=1} x_i!] \text{ (where } \Sigma^k_{i=1}=n) \quad \text{(Eq. 3)}$$

Thus, the probability that each of the k alleles classes will be represented at least once in the sample is given by

$$P(a_1>0, \ldots, a_k>0)= 1 - \{\Sigma^k_{i=1} P(a_i) - \Sigma^k_{1\le i<j\le k} P(a_i a_j) + \Sigma^k_{1\le i<j<z\le k} P(a_i a_j a_z) - \ldots (-1)^{k+1} \Sigma^k_{1\le i<j<z\ldots\le k-1} P(a_i a_j \ldots a_{k-1})\} \quad \text{(Eq. 4)}$$

(Crossa, 1989), where $P(a_i)$, the probability that the allele a_i will not appear in the sample, is $(1-p_i)^n$. Crossa (1989) evaluated this equation for the case of two, three and four alleles at different frequencies, whereas Marshall and Brown (1975) evaluated it for two and four alleles.

For m independent loci, the probability that each of k alleles classes will be detected at least once in each locus in a sample of size n is

$$\Pi^{m}_{l=1}\,[P(a_1>0,\ \ldots,\ a_k>0)] = \Pi^{m}_{l=1}[1 - \{\Sigma^{k}_{i=1}\,P(a_i) - \Sigma^{k}_{1\le i<j\le k}\,P(a_i a_j) + \Sigma^{k}_{1\le i<j<z\le k}\,P(a_i a_j a_z) - \ldots(-1)^{k+1}\Sigma^{k}_{1\le i<j<z\ldots\le k-1}\,P(a_i a_j \ldots a_{k-1})\}] \quad \text{(Eq. 5)}.$$

Although these equations can be evaluated by numerical procedures, for any number of loci and alleles at any frequency, obtaining the required sample size for many alleles at different frequencies in various loci is very impractical. Hernandez and Crossa (1992) developed a computer algorithm to evaluate these equations and therefore facilitate determining the optimal sample size. The authors specified the assumptions underlying Eq. 5 as: 1) seeds are sampled without regard to the genotype of the parents; 2) there are no associations among genes from different loci (linkage equilibrium); 3) if there are no associations between genes within individuals at any locus, then the required sample size is exactly half the sample size (n) given by Eq. 5; 4) if there is a perfect association between genes within individuals at any locus, then the required sample size equals the sample size (n) given by Eq.; and 5) if the degree of association between genes within individuals is unknown, then the required sample size is between n/2 and 2.

If frequencies of alleles are unknown ~~However~~, a more general equation for estimating an optimal sample size that will still retain at least one copy

of each of the k allele classes with a given probability is required. This can be obtained by assuming that k-1 alleles occur at an identical low frequency of p_o and that the k^{th} allele occurs at a frequency of $1-[(k-1)p_o]$. Then, Eq. 4 can be reduced to the following much simpler expression

$$P(a_1>0, \ldots, a_k>0) = 1 - \{\Sigma^{k-1}_{r=1}(-1)^{r-1} \binom{k-1}{r} (1-rp_o)^n\}$$

(Eq. 6) where r denotes the number of terms in the summation (see Appendix A).

By considering only the first term of Eq. 6 [i.e., $(k-1)(1-p_o)^n$], log transforming and solving for n, the resulting equation is

$$n > [\log(1-P)-\log (k-1)]/[\log(1-p_o)] \quad \text{(Eq. 7)}$$

It can be shown, for this case, that the other terms of Eq. 6 are negligible (see Appendix B). This general expression shows that the sample size required to retain, with probability P, at least one copy of each of the k allele classes at one locus depends on the number and the frequency of the alleles. A. H. D. Brown in Frankel and Soule (1981) evaluated this formula only for the case of two alleles per locus [i.e., $\log(k-1)=0$].

For the case of m independent loci and the same number of allele classes (k) at each locus, Eq. 7 can be written as

$$n > \{\log[1-(P)^{1/m}]-\log (k-1)\}/\log(1-p_o) \quad \text{(Eq. 8)}$$

The formula used by Chapman (1984) resembles Eq. 8, but the former considered only two alleles per locus.

In general, these formulae suggest that the optimal sample size is much more strongly affected by the frequency

of the rare alleles than it is by the number of alleles or by the number of loci. For example, for loci with two, three, or four alleles, each at a frequency of 0.05 (p_o), 89 to 110 additional individuals are required to retain at least one allele in each class at each of the 10 loci with 90% probability (Table 1); if 100 loci are involved, 45 more individuals are required in the sample. For two, three, or four alleles, ~~when~~ then the frequency of a particular allele declines from 0.05 to 0.03 in each of 10 loci, the sample size required to include at least one of the alleles from each class at each locus with 90% probability increases by 61 to 76 additional individuals (Table 1). Similarly, for 100 loci the sample size increases by 91 to 106 individuals. For loci with two, three, or four alleles, each at a frequency of 0.05 (p_o), 134 to 156 individuals are required to retain at least one allele in each class at each of the 50 loci with 95% probability (Table 2); if 150 loci are involved, 22 more individuals are required in the sample. To capture, with 90%-95% probability, at least one allele at 0.03 frequency in each allele class at each of 150 loci, sample sizes of 238-350 individuals are required. For preserving alleles at 0.01 frequency with 90%-95% probability, the regenerating sample size should include between 722 to 1057 plants.

These results indicate that sample sizes of 160-210 plants are required to capture alleles at frequencies of 0.05 or higher in each of 150 loci with 90-95% probability.

This sample size will restrict inbreeding to 1% per regeneration, and thereby avoid inbreeding depression. For most quantitative traits, alleles rarer than 0.05 would probably contribute little to the mean or the variance of the character in the population and so could not easily be measured or evaluated. Therefore, a 0.05 allelic frequency seems to be an appropriate level for calculating requisite sample sizes.

Multinomial models used here allow the generalization of variuos equations to k, rather than two, alleles. Although isozyme analysis in general resolved two or less alleles per locus, modern molecular markers techniques such as RFLP and RAPD analyses are often uncovering greater number of alleles per locus.

The probability of losing one or several alleles at a single locus when diploid individuals are sampled has been considered by Gregorius (1980), who compiled a table of the minimum sample sizes required to assure that all alleles with given frequencies are included with a certain probability. When the allelic frequency decreases from 0.05 to 0.03, or from 0.02 to 0.01, the required sample size approximately doubles.

Sample size of 90 individuals were suggested by Namkoong (1988) for an average loss of one allele at frequency of 0.05 at any of 100 loci. The required size increases to 458 individuals when the allelic frequency drops to 0.01.

THE CONCEPT OF CORE SUBSET

The ever increasing number and size of collections stored in germplasm banks, and the complexities of adequately managing and using them, have generated considerable concern within the world plant germplasm community.

The core subset concept was originally developed and described by Frankel and Brown (1984) and by Brown (1989a,b). Core subsets comprise specific accessions from an existing collection and therefore, do not constitute a separate collection *per se*. As such, they should be fully integrated with the "reserve subset" so that the collection is curated as an essential whole.

The aim of defining core subsets is the use of plant germplasm collections and provide efficient access to the range of genetic variation in the whole collection, making preliminary germplasm evaluations for needed traits more efficient.

Two of the most important questions that a germplasm manager must decide when forming core subsets are: 1) what is the optimal number of accessions that will include, with high probability, most of the alleles present in a given collection, and 2) how to select the accessions for the core subset.

Brown (1989a) considered four classes of alleles in a given germplasm collection: 1) common, localized, 2) rare, localized, 3) common, widespread, and 4) rare widespread. For the first three classes of alleles, Brown (1989a) studied the distribution of neutral alleles and obtained the expected number of alleles retained in a given size of the core subset. He recommended including 5% to 10% of the total collection and at least 3000 accessions per species. The last class consists of alleles that are always rare but that occur across most accessions of one collection. For the last two classes of alleles, each accession can be considered a random sample from the collection.

Determining the optimal number of accessions for a core subset. Assuring that widespread, common or rare alleles are captured

In this part of the paper, we are concerned with the minimal number of accessions that will retain most of the alleles that generally occur at low frequencies (less than 0.10) in most accessions of the collection. For this case, the expected number of alleles (n_A) in a sample of n accessions can be estimated as follows: Consider A alleles at a locus with frequencies of p_1, p_2, p_3, ..., p_j ,...., p_A across all accessions. Let the event $x_j=1$, if the j^{th} allele is included in a sample of n accessions and $x_j=0$ if not. The probability that the j^{th} allele is absent from the sample is

$P(x_j=0)=(1-p_j)^n$ and the probability of the j^{th} allele occuring in the sample is $P(x_j=1)=1-P(x_j=0)=1-(1-p_j)^n$. Then, let $S=x_1+x_2+\ldots+x_A$ ($S=\Sigma_{j=1}\ x_j$, where $j=1,2,\ldots,A$) be the number of alleles per locus present in the sample of n accessions. The expected value of x_j is given by $E(x_j)=\Sigma_{xj=0}\ x_j\ P(x_j)=(0)(1-p_j)^n+(1)[1-(1-p_j)^n] = 1-(1-p_j)^n$. Thus, the expected number of alleles (n_A) per locus in a sample of n accessions is $E(S)= n_A = E(\Sigma^A_{j=1}\ x_j) = \Sigma^A_{j=1}E(x_j) = \Sigma^A_{j=1}\ [1-(1-p_j)^n] = A-\Sigma^A_{j=1}\ (1-p_j)^n$. The expected number of alleles captured in all loci is the sum of n_A for each locus (Brown, 1989a)

Some numerical examples employing this formula are presented in Table 3. When four of five alleles have frequencies of 0.05 and one has a frequency of 0.80, a subset of 25 accessions would be expected to include 4 out of 5 alleles. When two of three alleles have frequencies of 0.05 and one has a frequency of 0.90, a subset of 15 accessions would retain, on the average, two of the three original alleles. For rare alleles widespread throughout the collection and the range of allelic frequencies considered here, most of the alleles with frequencies of 0.03 and 0.05 per locus will be included in a subset of 25-100 accessions. These results apply only to alleles that are widespread throughout the collection.

In general, a useful strategy for forming core subsets in maize would be to use a stratified sampling strategy. For

example, subdividing the total number of accessions into non-overlapping groups based on racial complex and/or ecogeographical criteria. Then, within each racial complex select 25 to 100 accessions. A subset of this size will preserve, on the average, alleles with frequencies higher than 0.03 in each race collection. If possible, races represented by fewer accessions in the collection should be collected more throughly to sample alleles with frequencies higher than 0.03. Within each race complex accessions can be grouped by region or elevation.

Accessions should be placed in multilocational replicated trials and several morph-agronomic attributes should be measured. Classification techniques such as cluster analysis and ordination methods such as principal components analysis have proved to be useful for assessing genetic diversity and therefore help the curator to identify very similar accessions within racial or ecogeographical subgroups of the core (Crossa et al., 1992).

In CIMMYT, a core subset for the Tuxpeño race collection was formed from 848 original accessions. Of these, 175 were selected based on lodging and general adaptation and placed in a replicated trial at two locations (Crossa et al., 1992). These accessions were collected in two different ecogeographical regions, dry ecology and wet ecology. A total of 40 accessions were selected to form the final core subset.

APPENDIX A

The term $\Sigma^{k}_{i=1} P(a_i)$ of Eq. 4 includes k-1 sub-terms without the allele a_k and one sub-term with a_k. The probability that any of the k-1 alleles will be absent from the sample is $(1-p_o)^n$ and the probability that the k^{th} allele be absent from the sample is $[(k-1)p_o]^n$. Thus, the term $\Sigma^{k}_{i=1} P(a_i)$ can be written as

$$(k-1)(1-p_o)^n + [(k-1)p_o]^n$$

The term $\Sigma^{k}_{1\leq i<j\leq k} P(a_i a_j)$ of Eq. 4 includes $\binom{k-1}{1}$ sub-terms that contain the allele a_k, each with a probability of $[(k-2)p_o]^n$, and $\binom{k-1}{2}$ sub-terms that do not include the allele a_k, each with a probability of $(1-2p_o)^n$. Therefore, the term $\Sigma^{k}_{1\leq i<j\leq k} P(a_i a_j)$ can be summarized as follows

$$\binom{k-1}{1}[(k-2)p_o]^n + \binom{k-1}{2}(1-2p_o)^n$$

The term $\Sigma^{k}_{1\leq i<j<z\leq k} P(a_i a_j a_z)$ of Eq. 4 comprises $\binom{k-1}{2}$ sub-terms that contain the allele a_k, each with a probability $[(k-3)p_o]^n$, and $\binom{k-1}{3}$ sub-terms that do not include the allele a_k, each with a probability of $(1-3p_o)^n$. Then, the $\Sigma^{k}_{1\leq i<j<z\leq k} P(a_i a_j a_z)$ term is reduced to

$$\binom{k-1}{2}[(k-3)p_o]^n + \binom{k-1}{3}(1-3p_o)^n$$

So, in general, the r^{th} term of Eq. 4 can be expressed as follows

$$\binom{k-1}{r-1}[(k-r)p_o]^n + \binom{k-1}{r}(1-rp_o)^n$$

Therefore, Eq. 4 is reduced to

$$P(a_1>0, \ldots, a_k>0) = 1 - \{\Sigma^{k-1}_{r=1}(-1)^{r-1}[\binom{k-1}{r-1}[(k-r)p_o]^n + \binom{k-1}{r}(1-rp_o)^n\}$$

Since the term $[(k-r)p_o]^n$ is so small that is negligible,

$$P(a_1>0, \ldots, a_k>0) = 1 - \{\Sigma^{k-1}_{r=1}(-1)^{r-1} \binom{k-1}{r} (1-rp_o)^n\}$$

APPENDIX B

We can substitute the value of n from Eq. 7 into any of the other summation term of Eq. 6 and prove that the value of that term is so small that is negligible. The r^{th} term of the summation can be written as $\binom{k-1}{r} (1-rp_o)^a$ (for $r \geq 2$), where $a=n=[\log(1-P)-\log(k-1)]/\log(1-p_o)$ from Eq. 7. That is, $(1-rp_o)^a=\exp[\{(\log(1-P)-\log(k-1))/\log(1-p_o)\}\{\log(1-rp_o)\}]$ Because $\{[\log(1-P)-\log(k-1)]/\log(1-p_o)\}\{\log(1-rp_o)\}]$ is negative, the maximum value of $(1-rp_o)^a$ occurred when the expression $\exp[\{(\log(1-P)-\log(k-1))/\log(1-p_o)\}\{\log(1-rp_o)\}]=1$. Therefore, minimum values of p_o, k, and r that make the quantity $\{[\log(1-P)-\log(k-1)]/\log(1-p_o)\}\{\log(1-rp_o)\}$ approach to 0 are required. When $po \to 0$, the limit of $\log(1-rp_o)/\log(1-p_o)$ approaches r. Then $(1-rp_o)^a=\exp\{[\log(1-P)-\log(k-1)](r)\}=\exp[(r)\log\{(1-P)/(k-1)\}]=[(1-P)/(k-1)]^r$ which is minimized when r=2 and k=3. Therefore, the maximum value of $(1-rp_o)^a$ is at $[(1-P)/2]^2$ and for P=0.9 and 0.95, $(1-rp_o)^a=0.0025$ and 0.000625, respectively. These are the maximum possible values that the second summation term of Eq. 6 can take.

Table 1. Sample sizes required to achieve a 90% probability of including at least one copy of alleles with p_o of 0.05, 0.03, and 0.01 from each allele class for several alleles at each locus.

Number of alleles	Number of loci: 1	2	5	10	50	100	150
	---------- p_o=0.05 ----------						
2	45	58	75	89	120	134	142
3	58	71	89	102	134	147	155
4	66	79	97	110	142	155	163
10	88	101	118	132	163	177	184
15	96	109	127	140	172	185	193
	---------- p_o=0.03 ----------						
2	76	97	127	150	202	225	238
3	98	120	150	172	225	248	261
4	112	134	163	186	238	261	274
10	148	170	199	222	274	297	311
15	162	184	214	236	289	312	325
	---------- p_o=0.01 ----------						
2	229	295	385	454	613	682	722
3	298	364	454	523	682	751	791
4	338	405	494	563	723	791	832
10	448	514	604	672	832	901	941
15	492	558	648	716	876	945	985

Table 2. Sample sizes required to achieve a 95% probability of including at least one copy of alleles with p_o of 0.05, 0.03, and 0.01 from each allele class for several alleles at each locus.

Number of Alleles	Number of loci: 1	2	5	10	50	100	150
				p_o=0.05			
2	58	72	89	103	134	148	156
3	72	85	103	116	148	161	169
4	80	93	111	124	156	169	177
10	101	115	132	146	177	191	198
15	110	123	141	154	186	199	207
				p_o=0.03			
2	98	121	151	173	226	249	262
3	121	143	173	196	249	271	285
4	134	157	187	209	262	285	298
10	170	193	223	245	298	321	334
15	185	207	237	260	313	335	349
				p_o=0.01			
2	298	366	456	525	685	754	794
3	367	435	525	594	754	823	863
4	407	475	565	634	794	863	903
10	517	584	675	744	903	972	1013
15	561	628	719	787	947	1016	1057

Table 3. Number of accessions (n) required for a core subsets so that n_A alleles per locus are retained for loci with 3, 4 and 5 alleles at different frequencies.

Allelic frequency						
p_1	p_2	p_3	p_4	p_5	n	n_A
0.0001	0.0001	0.9998			2000	1
0.01	0.01	0.98			70	2
0.01	0.01	0.98			300	3
0.03	0.03	0.94			90	3
0.05	0.05	0.90			15	2
0.0001	0.0001	0.0001	0.9997		4000	2
0.01	0.01	0.01	0.97		100	3
0.01	0.01	0.01	0.97		350	4
0.03	0.03	0.03	0.91		100	4
0.05	0.05	0.05	0.85		25	3
0.0001	0.0001	0.0001	0.0001	0.9996	6000	3
0.01	0.01	0.01	0.01	0.96	150	4
0.01	0.01	0.01	0.01	0.96	400	5
0.03	0.03	0.03	0.03	0.88	100	5
0.05	0.05	0.05	0.05	0.80	25	4

REFERENCES

Brown, A.H.D. 1989a. The case for core collections. In: Brown, A. H. D. Brown, O. H. Frankel, D. R. Marshall and J. T. Williams (eds.), The use of plant genetic resources. Cambridge University Press, Cambridge, UK. pp. 136-156.

Brown, A.H.D. 1989b. Core collections: A practical approach to genetic resources management. Genome 31:818-824.

Chapman, C.G.D. 1984. On the size of a genebank and the genetic variation it contains. In: Holden J. H. W., J. T. Williams (eds.), Crop genetic resources: Conservation and evaluation. Allen and Unwin, London, UK. pp 102-108.

Crossa, J. 1989. Methodologies for estimating the sample size required for genetic conservation of outbreeding crops. Theor. Appl. Genet. 77:153-161.

Crossa, J., S. Taba, S. Eberhart, P. Bretting. 1992. Practical methods for maintaining germplasm of outbreeding crops. Crop Science (submitted).

Frankel, O.H., and M.E. Soule. 1981. Conservation and evolution. Cambridge University Press, Cambridge.

Frankel, O.H., and A.H.D. Brown. 1984. Plant genetic resources today: A critical appraisal. In: Holden J. H. W., J. T. Williams (eds.), Crop genetic resources: Conservation and evaluation. Allen and Unwin, London, UK. pp. 149-257.

Gregorius, H.R. 1980. The probability of losing an allele when diploid genotypes are sampled. Biometrics 36:643-652.

Hernandez C.M. and J. Crossa. 1992. A program for estimating the optimum sample size for germplasm conservation. J. of Heredity (in press).

Marshall, D.R., and A.H.D. Brown. 1975. Optimum sampling strategies in genetic conservation. In Frankel, O. H., J. G. Hawkes (eds.), Crop genetic resources for today and tomorrow. Cambridge University Press, Cambridge, UK. pp. 53-80.

Namkoong, G. 1988. Sampling for germplasm collection. Hort. Science 23: 79-81.

Weir, B.S. 1990. Sampling properties of gene diversity. In A.H.D. Brown, M.T. Clegg, A.L. Kahler, B.S. Weir (eds), Plant population genetics, breeding, and genetic resources. Sinauer Associates, Inc., Sunderland, pp. 23-42.